AD-771 575

GROUP JUDGMENT TECHNOLOGY

Norman Dalkey

California University

Prepared for:

Air Force Systems Command
Advanced Research Projects Agency

30 June 1973

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD-771575

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
GROUP JUDGEMENT TECHNOLOGY

5. TYPE OF REPORT & PERIOD COVERED
(Interim) Semi-Annual Technical - 1Jan-30Jun73

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Dr. Norman Dalkey

8. CONTRACT OR GRANT NUMBER(s)
F30602-72-C-0429

9. PERFORMING ORGANIZATION NAME AND ADDRESS
The Regents of the University of California
Dept of Psychology
405 Hilgard Avenue, Los Angeles, CA 90024

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
61101E
ARPA 22080101

11. CONTROLLING OFFICE NAME AND ADDRESS
Rome Air Development Center (IRDA)
Griffiss AFB NY 13441

12. REPORT DATE
30 June 1973

13. NUMBER OF PAGES
~~147~~ 179

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
N/A

16. DISTRIBUTION STATEMENT (of this Report)
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
5.4.1.1; 5.10.2.1; 12.1.1.2.1; 15.7.8.4.1; 15.7.8.8.1

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Delphi Project is to develop improved procedures for the use of expert judgement in decision-making through computer-based on-line decision structure which can be employed to formulate policy problems and select preferred policy options. The laboratory configuration was completed and a series of experiments were conducted. The next progress report will be published 1 Jan 74.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

GROUP JUDGEMENT TECHNOLOGY

Dr. Norman Dalkey

Contractor: University of California at Los Angeles
Contract Number: F30602-72-C-0429
Effective Date of Contract: 1 July 1972
Contract Expiration Date: Merged with F30602-74-C-0016 was 29 June 1975
Amount of Contract: $666,441.00
Program Code Number: 3D20

Principal Investigator: Dr. Norman C. Dalkey
Phone: 213 825-4315

Project Engineer: Patricia M. Langendorf
Phone: 315 330-3626

This research was supported by the Defense Advanced Research Projects Agency of the Department of Defense and was monitored by Patricia M. Langendorf, RADC (IRDA), GAFB, NY 13441 under contract F30602-72-C-0429.

TITLE: GROUP JUDGEMENT TECHNOLOGY (U)

PUBLICATION REVIEW

This technical report has been reviewed and is approved.

RADC Project Engineer

JAMES D. EDGINGTON, Lt Col, USAF
Asst Ch, Intel Data Handling Branch
Intel and Recon Division

ABSTRACT

The Delphi Project is to develop improved procedures for the use of expert judgement in decision-making through computer-based on-line decision structure which can be employed to formulate policy problems and select preferred policy options. The laboratory configuration was completed and a series of experiments were conducted. The next progress report will be published 1 Jan 74.

PROGRESS ON PROJECT: 1 JANUARY 1973 THROUGH 30 JUNE 1973

Aims

The Delphi Project has as its continuing goal the development of improved procedures for the use of expert judgment in decision-making. The long-range program to implement this aim in the CCBS context is the development of a prototype on-line decision structure which can be employed by governmental agencies or industrial decision-making groups to formulate policy problems and select preferred policy options.

The orientation of the project is primarily experimental and technological, i.e., the laboratory testing of promising innovations in group judgment procedures. In addition, about 20% of the effort is expended in developing related analytical tools and in pursuing theoretical developments which can further the experimental program.

Activities through June 1973

Laboratory Experiments

The laboratory configuration for running Delphi experiments was completed during this period, and a series of experiments was conducted, using upper-class and graduate students at UCLA as subjects. The experiments were concerned with the utilization

of additional (factual) information by the subjects in revising their estimates of almanac type questions.

The practical import of this research arises from the fact that in traditional procedures for applying group judgment to governmental and industrial decision making (panels, committees, staff studies, etc.) a major part of the activity of the group is taken up by assimilating information which is not "in the heads" of the group when it is convened. The assimilative activity involves collecting and reading additional books and reports, site visits, briefings by other experts, and the like. Most of these activities are not compatible with Delphi procedures. Geographic dispersal of panels and use of mailed questionnaires precludes transmitting large amounts of information to each member of the group; for many Delphi studies, the panels are large (50 or more members) which rules out many of the traditional information handling procedures such as face-to-face discussion; and generally, Delphi studies encompass a wide range of subject matter which makes the relevant additional material unmanageably large. It has been an underlying assumption of many Delphi practioners that using a large group of expert respondents assures that most of the relevant information is already in the heads of the respondents, or readily available to them as individuals.

There are reasons for suspecting that this assumption is not justified. It is clear from the amount of effort assigned to

self-education by committees that this activity is considered to be one of the most important components of the group judgment process. In addition, the rather sparse experimental data that exists on the subject indicates that utilization of additional factual information can increase the accuracy of group judgment by a large factor.

The development of procedures for dealing with additional factual information which can approximate the desirable features of the traditional "self-education" activities of committees, and at the same time can be utilized in a Delphi context is thus one of the crucial areas where additional capabilities are needed. The experimental series began this spring is aimed at obtaining data which will be useful for designing and testing such techniques.

In earlier experiments conducted at RAND, feed-in of additional factual material was identified as one of the most effective procedures for improving both individual and group responses.* When facts were fed in in random order, on the average there was a decreasing return for each additional fact. In the present

---

*RM-6118-PR and Dalkey, Rourke, Lewis and Snyder, Studies in the Quality of Life, D.C. Heath, 1972 pp. 48-54.

series of experiments, subjects are presented with a menu of four potential facts (that is, the facts are described but numerical values are not stated.). The subjects rate each of the four in terms of their expected value in improving the answer to the primary question. The facts are then presented in one of two orders; individually to each subject in the order of his rating, or collectively to all subjects in the order of the group average rating.

To date, four experimental sessions involving 10 subjects per session and 10 primary questions, have been run in the individual selection mode, and five sessions varying from five to 10 subjects, and four to 10 questions have been run in the group selection mode. Preliminary analysis of the data indicates that the hypothesis that error will decrease much more rapidly under individual selection of facts than under random presentation is borne out. In addition, the decrease in return per additional fact is greater with self-selection.

First returns on the group selection mode indicate that the decrease in error is even more rapid under group selection; however, the data is less extensive than for the individual mode, and appears to be interacting with the feedback of medians and modes of the first round answers. (In the RAND experiments, statistical feedback in conjunction with additional facts did not appear to have a significant effect on group error. In the present experiment, it appears that it might. This could

indicate the existence of feedback, feed-in interaction in the case that the facts are selected, rather than being presented at random.)

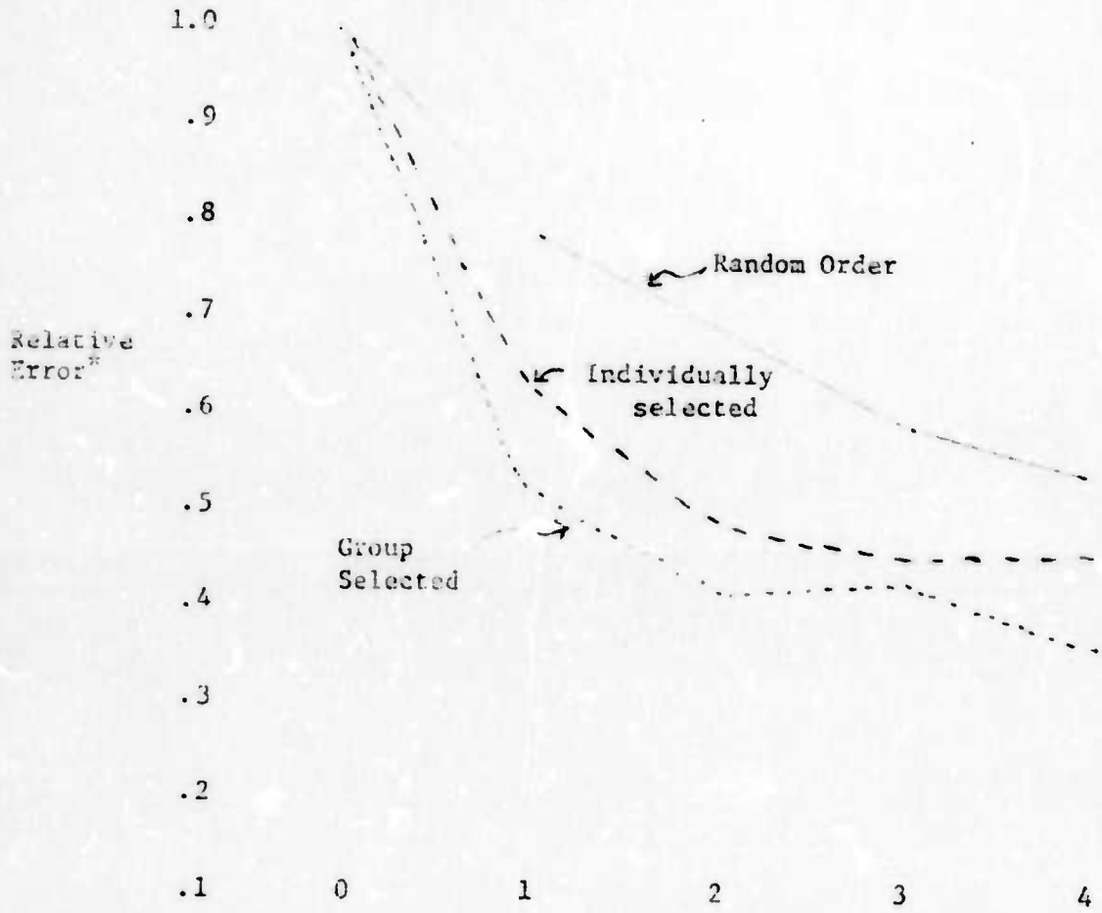
The accompanying figure shows the preliminary results concerning feed-in of facts. The upper curve shows the results for random feed-in for comparison.

In experimental sessions, the CCBS interactive system has worked smoothly and response time of the system has been well within the tolerance level of subjects. The interactive features of the system have been crucial in conducting the group selection mode of the feed-in of additional material, and hence, the experiments have shown the value of the system for conducting more complex types of Delphi exercises.

One unanticipated effect of the present configuration is the relatively slow pace of the group in completing a cycle of responses to a given question. This appears to be almost entirely a result of the large variation in individual response times for the various subtasks. It is possible that the complete isolation of individual subjects in separate rooms contributes to this variability, and subsequent experiments will try more "open" configurations to allow a measure of group self-pacing. It is hoped that this tactic will increase the amount of data collected within a given experimental session.

SUMMARY

ADDITIONAL FACTS



*Error on round one (no facts) = 1.

in addition to replicating some of the experiments concerned with feed-in of relevant facts (primarily to increase the amount of data to generate an acceptable level of significance) similar experiments will be run during the fall where the mode of response will be a probability distribution, rather than a point estimate. These experiments will furnish data to assess the reliability of probability estimates, to evaluate the effect of factual feed-in for probability estimates, and to test several hypotheses concerning the dependence of anonymous probability judgments.

Although at the moment the data is not sufficient to allow finalizing techniques for integrating external information in an applied Delphi exercise, a "first approximation" appears to be emerging from the experimental results. A summary of this hypothetical set of procedures is perhaps the clearest way to show the practical import of the relevant fact experiments: (1) The first judgmental round of a Delphi exercise should be run in the "standard" way -- i.e., anonymous judgments from the panel, using only the information in their heads, or what is conveniently incorporated into the questionnaire itself. (2) The first round questionnaire should include an open-ended set of responses, where each expert is asked to list information which he thinks is available, feasible to obtain in a reasonable time, and is important for increasing the certainty of the group judgment. (3) For those questions where there is significant

disagreement on the first round (significant in the statistical sense) and where the list of desired information from the respondents is highly diverse, an auxiliary round is run to obtain judgments of relative importance of the nominated information. (4) Questions with high agreement on the first round are not iterated. Questions with disagreement on the answer, but reasonable agreement on the needed information, are iterated, with as much of the nominated information as can be obtained included. For questions with disagreement on both answers and desired information, the results of the auxiliary judgments concerning importance and feasibility are used to select the items of information which are researched and fed into the groups. (5) For those cases where the nominated information is relatively extensive, a division of labor is designed, based on the individual judgments of importance, where the relevant information is apportioned among the members of the group in roughly equal proportions.

During the research (collection) phase of this procedure, a quick round may be intercalated to determine if one or more members of the group know the specific information requested, or can obtain it easily. Here the standards for "know" would be very high.

This structure has not yet been tested. It will be tested in several exercises, beginning in the fall of 1973.

Utilization of TRACE for Data Analysis

An exercise is under way to assess the potential role of the TRACE system for analysis of the data base being generated by the Delphi experiments. The TRACE data manipulation capabilities are well suited for the complex and multiply nested data structures produced by Delphi exercises. Development of a convenient way to routinely process the data by TRACE should increase the speed and depth of analysis by a large factor. This becomes especially important in studying the "fine-grain" effects of order of facts, fact ratings, and the like.

Design of Formatted Conference

The initial design has been completed for a software system that will allow a geographically dispersed conference (e.g., via the ARPA Network) to be conducted on a variety of decision problems. The software system is described as a formatted conference, to distinguish it from a number of other computer conferencing systems that operate in a more open-ended fashion (e.g., the conferencing system developed at OEP by Murray Turoff.)

The system embodies a highly general model of the problem identification and group decision process: specification of objectives or goals, specification of policy alternatives, and judgments of the expected contribution of alternatives to objectives. The major methodological problem facing such a

general group process is the formulation of techniques for reducing the highly diverse attitudes and points of view of different members of a panel to a coherent and manageable structure. In the prototype decision system under development, major reliance will be put on dimension-reduction techniques such as multi-dimensional analysis, cluster analysis and factor analysis. As of the moment, it looks as if cluster analysis will be the most useful of these techniques, but the system will incorporate the capability to use all three, thus allowing comparison of their relative value in the shakedown exercises.

Initially, it is being assumed that identification of policy alternatives can be carried out in much the same way as identification of goals and objectives; thus a common group process is being set up for these two activities. The process consists of the following basic steps: (1) Listing by each panel member of the most important goals and objectives (or policy alternatives) for the present decision problem as he perceives them. (2) Formulation of a master list of these individual items. (3) Group editing of the items in the master list, including ratings for clarity, for relevance (or importance) and for identity among items. (4) Withdrawal or rewriting of unclear items, and consolidation of identical items. (5) Sorting by each individual of the edited list into what he perceives as similar categories. (6) Use of a cluster routine (or multi-dimensional scaling or factor analytic routine) to aggregate the individual classification and formulate a (much shorter) list of aggregated goals and

objectives (or policy alternatives).

The software system should be running in prototype in the Fall of 1973. Shakedown exercises will investigate both the efficiency of the system as a method of decision formulation and the meaningfulness of the aggregated objectives and policy alternatives to the group members. By the spring of 1974, the system should by ready for a pilot study with either a governmental agency group, or an industrial contractor. The selection of an appropriate client will be made in cooperation with ARPA.

Other uses of the decision analysis system are envisaged in addition to direct use in decision exercises by operating agencies. The output of the system can be used in a number of experimental studies of basic aspects of group decision-making. One of the important features of cluster analysis routines is that they generate a hierarchy of items at different levels of generality. Put shortly, the system will generate a set of nested decision models of various levels of abstractness. Such a set of models will enable the testing of hypotheses concerning the optimal level of detail for formulating decision problems, as a function of the degree of uncertainty of the group concerning basic elements of the problem. Similarly, they will enable testing of various hypotheses concerning the optimal way to combine indices of uncertainty (dispersions or self ratings) for complex judgments.

Analytic Developments

Cross-Impact matrices. In the previous semi-annual report, a basic development in the theory of cross-impact matrices was described. This is formulation of a distance measure whereby a set of probabilities can be mapped onto a metric space. Since then a beginning has been made in treating the theory of cross-impacts as a function of time. The elementary theory has been fully developed, but is too complex for event sets of the size met with in practice. Work is continuing on the derivation of a reasonable approximation for large sets of events.

Probability Aggregation. The analysis of group judgment defined as an aggregation of probability judgments by individuals has been extended by application of probabilistic scoring rules to several potential aggregation methods. The major outcome of this analysis is that if the aggregation rule is a statistical composite (a mean, geometric mean, or the like) then the probabilistic score of the composite is always greater than or equal to the average probabilistic score of the individual members; if the aggregation rule is an optimal estimate, employing Bayes' rule or an inverse Bayes' rule, then the net probabilistic score (defined as the score, minus the score that would be obtained by using the a priori probabilities as an estimate) is given by the expression $S(G) = n\overline{S(I)} + D$, where $S(G)$ is the net score of the group, $\overline{S(I)}$ is the average net score of the individual members, n is the number of members of

the group, and D is the expected dependence among the individual estimates. If D is small or positive, and the average member of the group has a positive net score, then the group does "n times as well" as the average individual; if on the other hand, the average net score of the individual members is negative, then the group does "n times as bad" as the average individual.

These results have been incorporated in a revised version of TM-37 "Delphi, Some Basic Considerations". Some of the upcoming experiments with probabilistic judgments will be used to generate data that will identify conditions under which the optimstic type of result can be expected to occur with probabilistic aggregation.

Model of Estimate Change with Feedback. A mathematical model has been developed to describe the change in group distributions of point estimates between round one and round two, when the group receives feedback of medians and quartiles from round one. Analysis of data on individual estimate change indicates that a good approximation is the individual shifts toward the median a constant fraction of his distance from the median on the first round; i.e., for a distribution where the median = 1,

$$x' = x(1-s) + s$$

where x' is the second round estimate, x is the first round estimate, and $0 \leq s \leq 1$ is the fractional change. From extensive data on individual changes, $s \sim .6$; i.e., the individual does a little more than "split the difference" between his location

and that of the median.

The second part of the model is the assumption that the first round distribution is log normal. This assumption has been well verified by previous studies. The geometric mean on the second round (GM') would be predicted to be (for a log normal distribution with mean of the log transform = 0 and standard deviation = 1.)

$$\ln GM' = \int_{s}^{\infty} \frac{\ln y}{\sqrt{2\pi(y-s)}} \; e^{-\left(\frac{\ln \frac{y-s}{1-s}}{2}\right)^2} dy$$

For s = .6, the equation predicts an increase of 11% in the geometric mean between round one and round two. This is slightly smaller than the observed increase of about 15%. At the moment it is not clear whether this discrepancy is due to departures from the assumption of lognormality on the part of specific distributions for specific questions, or whether there is some additional factor leading to improvement on feedback. The model ignores some of the change behavior in the vicinity of the median.

Additional Activities

Norman Dalkey participated in a UNESCO working conference on Man and the Biosphere at UNESCO Headquarters, Paris, France, during the week of March 26-30, 1973, where he presented

a paper on "State of the Art in Measurement of Quality of Life" and chaired the working group on methodology in research on environmental perception. He has acted as an informal consultant to the Public Health Service (NIH) on evaluation of research proposals, and to the Air Resources Board of the State of California on a Delphi evaluation of disability resulting from types and amounts of air pollutants. He also consulted with the Air Force Office of Medical Personnel on a service-wide Delphi study of medical personnel problems arising from the all-volunteer policy.

Forthcoming Reports

1. Elementary Cross Impact Model Including Dependence on Time.
2. A Model of Group Judgment Change as a Function of Statistical Feedback.

Reports Issued

TM-37 Delphi: Some Basic Considerations.

TM-42 A Delphi Study of Factors Affecting the Quality of Life.